PROBABILITY THEORY, AN ANALYTIC VIEW

DANIEL W. STROOCK
Massachusetts Institute of Technology

CAMBRIDGE
UNIVERSITY PRESS

PUBLISHED BY THE PRESS SYNDICATE OF THE UNIVERSITY OF CAMBRIDGE
The Pitt Building, Trumpington Street, Cambridge, United Kingdom

CAMBRIDGE UNIVERSITY PRESS
The Edinburgh Building, Cambridge CB2 2RU, UK http://www.cup.cam.ac.uk
40 West 20th Street, New York, NY 10011-4211, USA http://www.cup.org
10 Stamford Road, Oakleigh, Melbourne 3166, Australia
Ruiz de Alarcón 13, 28014 Madrid, Spain

First published 1993
Reprinted with corrections 1994
First paperback edition 1999

A catalog record for this book is available from the British Library

Library of Congress Cataloging in Publication data is available

ISBN 0 521 43123 9 hardback
ISBN 0 521 66349 0 paperback

Transferred to digital printing 2003

Contents

Chapter I:
Sums of Independent Random Variables

§1.1 Independence

In one way or another, most probabilistic analysis entails the study of large families of random variables. The key to such analysis is an understanding of the relations among the family members; and of all the possible ways in which members of a family can be related, by far the simplest is when the relationship does not exist at all! For this reason, we will begin by looking at families of *independent* random variables.

Let $(\Omega, \mathcal{F}, P)$ be a **probability space** (i.e., Ω is a nonempty set, $\mathcal{F}$ is a σ-algebra over Ω, and P is a measure on the measurable space $(\Omega, \mathcal{F})$ having total mass 1); and, for each i from the (nonempty) index set $\mathcal{I}$, let $\mathcal{F}_i$ be a sub σ-algebra of $\mathcal{F}$. We say that the σ-algebras $\mathcal{F}_i$, $i \in \mathcal{I}$, are **mutually** P**-independent** or, less precisely, P**-independent**, if, for every finite subset $\{i_1, \ldots, i_n\}$ of distinct elements of $\mathcal{I}$ and every choice of $A_{i_m} \in \mathcal{F}_{i_m}$, $1 \le m \le n$,

$$P\big(A_{i_1} \cap \cdots \cap A_{i_n}\big) = P\big(A_{i_1}\big) \cdots P\big(A_{i_n}\big). \tag{1.1.1}$$

In particular, if $\{A_i : i \in \mathcal{I}\}$ is a family of sets from $\mathcal{F}$, we say that A_i, $i \in \mathcal{I}$, are P**-independent** if the associated σ-algebras $\mathcal{F}_i = \{\emptyset, A_i, A_i\complement, \Omega\}$, $i \in \mathcal{I}$, are. To gain an appreciation for the intuition on which this definition is based, it is important to notice that independence of the pair A_1 and A_2 in the present sense is equivalent to

$$P\big(A_1 \cap A_2\big) = P\big(A_1\big)P\big(A_2\big),$$

the classical definition which one encounters in elementary treatments. Thus, the notion of independence just introduced is no more than a simple generalization of the classical notion of *independent pairs of sets* encountered in non-measure theoretic presentations; and therefore, the intuition which underlies the elementary notion applies equally well to the definition given here. (See Exercise 1.1.10 below for more information about the connection between the present definition and the classical one.)

As will become increasing evident as we proceed, infinite families of independent objects possess surprising and beautiful properties. In particular, mutually independent σ-algebras tend to *fill up space* in a sense which is made precise by the following beautiful thought experiment designed by A.N. Kolmogorov. Let $\mathcal{I}$ be any index set, take $\mathcal{F}_\emptyset = \{\emptyset, \Omega\}$, and for each nonempty subset $\Lambda \subseteq \mathfrak{I}$, let

$$\mathcal{F}_\Lambda = \bigvee_{i\in\Lambda} \mathcal{F}_i$$

be the σ-algebra generated by $\bigcup_{i\in\Lambda} \mathcal{F}_i$ (i.e., the smallest σ-algebra containing all of the $\mathcal{F}_i$'s). Next, define the **tail σ-algebra** $\mathcal{T}$ to be the intersection over finite $\Lambda \subseteq \mathcal{I}$ of the σ-algebras $\mathcal{F}_{\Lambda\complement}$. When $\mathcal{I}$ itself is finite, $\mathcal{T} = \{\emptyset, \Omega\}$ and is therefore ***P*-trivial** in the sense that $P(A) \in \{0,1\}$ for every $A \in \mathcal{T}$. The interesting remark made by Kolmogorov is that even when $\mathcal{I}$ is infinite, $\mathcal{T}$ is P-trivial whenever the original $\mathcal{F}_i$'s are P-independent. To see this, first note that, by assumption, $\mathcal{F}_{F_1}$ is P-independent of $\mathcal{F}_{F_2}$ whenever F_1 and F_2 are finite, disjoint subsets of $\mathcal{I}$. Since for any (finite or not) $\Lambda \subseteq \mathcal{I}$, $\mathcal{F}_\Lambda$ is generated by the algebra

$$\bigcup \{\mathcal{F}_F : F \text{ is a finite subset of } \Lambda\},$$

it follows (cf. Exercise 1.1.12) first that $\mathcal{F}_\Lambda$ is P-independent of $\mathcal{F}_{\Lambda\complement}$ for every $\Lambda \subseteq \mathcal{I}$ and then that $\mathcal{T}$ is P-independent of $\mathcal{F}_\mathcal{I}$. But $\mathcal{T} \subseteq \mathcal{F}_\mathcal{I}$, which means that $\mathcal{T}$ is *independent of itself*; that is, $P(A\cap B) = P(A)P(B)$ for all A, $B \in \mathcal{T}$. Hence, for every $A \in \mathcal{T}$, $P(A) = P(A)^2$, or, equivalently, $P(A) \in \{0,1\}$; and so we have now proved the following famous result.

1.1.2 Theorem (Kolmogorov's 0–1 Law). *Let $\{\mathcal{F}_i : i \in \mathcal{I}\}$ be a family of P-independent sub-σ-algebras of $(\Omega, \mathcal{F}, P)$, and define the tail σ-algebra $\mathcal{T}$ as above. Then, for every $A \in \mathcal{T}$, $P(A)$ is either* 0 *or* 1.

To get a feeling for the kind of conclusions which can be drawn from Kolmogorov's 0–1 Law (cf. Exercises 1.1.18 and 1.1.19 below as well), let $\{A_n\}_1^\infty$ be a sequence of subsets of Ω, and recall the notation

$$\text{(1.1.3)} \qquad \begin{aligned} \overline{\lim_{n\to\infty}} A_n &\equiv \bigcap_{m=1}^{\infty} \bigcup_{n\geq m} A_n \\ &= \big\{\omega : \omega \in A_n \text{ for infinitely many } n \in \mathbb{Z}^+\big\}. \end{aligned}$$

Obviously, $\overline{\lim}_{n\to\infty} A_n$ is measurable with respect to the tail field determined by the sequence of σ-algebras $\{\emptyset, A_n, A_n\complement, \Omega\}$, $n \in \mathbb{Z}^+$; and therefore, if the A_n's are P-independent elements of $\mathcal{F}$, then

$$P\Big(\overline{\lim_{n\to\infty}} A_n\Big) \in \{0,1\}.$$

In words, this conclusion can be summarized as the statement that: *for any sequence of P-independent events A_n, $n \in \mathbb{Z}^+$, either P-almost every $\omega \in \Omega$ is in infinitely many A_n's or P-almost every $\omega \in \Omega$ is in at most finitely many A_n's.* A more quantitative statement of this same fact is contained in the second part of the following useful result.

1.1.4 Borel–Cantelli Lemma. *Let $\{A_n : n \in \mathbb{Z}^+\} \subseteq \mathcal{F}$ be given. Then*

$$\sum_{n=1}^{\infty} P(A_n) < \infty \implies P\left(\varlimsup_{n\to\infty} A_n\right) = 0. \tag{1.1.5}$$

Conversely, if the A_n's are P-independent sets, then

$$\sum_{n=1}^{\infty} P(A_n) = \infty \implies P\left(\varlimsup_{n\to\infty} A_n\right) = 1. \tag{1.1.6}$$

(See part **(iii)** *of Exercise 5.2.34 and Lemma 8.5.46 for variations on this theme.)*

PROOF: The first assertion is an easy application of countable additivity. Namely, by countable additivity,

$$P\left(\varlimsup_{n\to\infty} A_n\right) = \lim_{m\to\infty} P\Bigl(\bigcup_{n\ge m} A_n\Bigr) \le \lim_{m\to\infty} \sum_{n\ge m} P(A_n) = 0$$

if $\sum_{n=1}^{\infty} P(A_n) < \infty$.

To prove (1.1.6), note that, by countable additivity, $P\left(\varlimsup_{n\to\infty} A_n\right) = 1$ if and only if

$$\lim_{m\to\infty} P\Bigl(\bigcap_{n\ge m} A_n\complement\Bigr) = P\left(\bigcup_{m=1}^{\infty}\bigcap_{n\ge m} A_n\complement\right) = P\Bigl(\Bigl(\varlimsup_{n\to\infty} A_n\Bigr)\complement\Bigr) = 0.$$

But, again by countable additivity, for given $m \ge 1$ we have that:

$$P\left(\bigcap_{n=m}^{\infty} A_n\complement\right) = \lim_{N\to\infty} \prod_{n=m}^{N} \bigl(1 - P(A_n)\bigr) \le \lim_{N\to\infty} \exp\left[-\sum_{n=m}^{N} P(A_n)\right] = 0$$

if $\sum_{n=1}^{\infty} P(A_n) = \infty$. (In the preceding, we have used the trivial inequality $1 - t \le e^{-t}$, $t \in [0,\infty)$.) $\square$

Another, and perhaps more dramatic, statement of the conclusion drawn in the second part of the preceding is the following. Let $\mathbf{N}(\omega) \in \mathbb{Z}^+ \cup \{\infty\}$ be the number of $n \in \mathbb{Z}^+$ such that $\omega \in A_n$. If the A_n's are independent, then Tonelli's Theorem implies that (1.1.6) is equivalent to[†]

$$P(\mathbf{N} < \infty) > 0 \implies \mathbb{E}^P[\mathbf{N}] < \infty.$$

[†] Throughout this book, we use $\mathbb{E}^P[X,\, A]$ to denote the expected value under P of X over the set A. That is, $\mathbb{E}^P[X,\, A] = \int_A X\, dP$. Finally, when $A = \Omega$ we will write $\mathbb{E}^P[X]$.

Having described what it means for the σ-algebras to be P-independent, we can now transfer the notion to random variables on $(\Omega, \mathcal{F}, P)$. Namely, for each $i \in \mathcal{I}$, let X_i be a **random variable** (i.e., a measurable function on $(\Omega, \mathcal{F})$) with values in the measurable space $(E_i, \mathcal{B}_i)$. We will say that the random variables X_i, $i \in \mathcal{I}$, are **(mutually) *P*-independent** if the σ-algebras

$$\sigma(X_i) = X_i^{-1}(\mathcal{B}_i) \equiv \{X_i^{-1}(B_i) : B_i \in \mathcal{B}_i\},\ i \in \mathcal{I},$$

are P-independent. Using

$$B(E; \mathbb{R}) = B\big((E, \mathcal{B}); \mathbb{R}\big)$$

to denote the space of bounded measurable $\mathbb{R}$-valued functions on the measurable space $(E, \mathcal{B})$, notice that P-independence of $\{X_i : i \in \mathcal{I}\}$ is equivalent to the statement that

$$\mathbb{E}^P\left[f_{i_1} \circ X_{i_1} \cdots f_{i_n} \circ X_{i_n}\right] = \mathbb{E}^P\left[f_{i_1} \circ X_{i_1}\right] \cdots \mathbb{E}^P\left[f_{i_n} \circ X_{i_n}\right] \tag{1.1.7}$$

for all finite subsets $\{i_1, \dots, i_n\}$ of distinct elements of $\mathcal{I}$ and all choices of $f_{i_1} \in B\big(E_{i_1}; \mathbb{R}\big), \dots,$ and $f_{i_n} \in B\big(E_{i_n}; \mathbb{R}\big)$. Finally, if we use $\mathbf{1}_A$ given by

$$\mathbf{1}_A(\omega) \equiv \begin{cases} 1 & \text{if} \quad \omega \in A \\ 0 & \text{if} \quad \omega \notin A \end{cases}$$

to denote the **indicator function** of the set $A \subseteq \Omega$, notice that the family of sets $\{A_i : i \in \mathcal{I}\} \subseteq \mathcal{F}$ is P-independent if and only if the random variables $\mathbf{1}_{A_i}$, $i \in \mathcal{I}$, are P-independent.

Thus far we have discussed only the abstract notion of independence and have yet to show that the concept is not vacuous. In the modern literature, the standard way to construct lots of independent quantities is to take products of probability spaces. Namely, if $\big(E_i, \mathcal{B}_i, \mu_i\big)$ is a probability space for each $i \in \mathcal{I}$, one sets $\Omega = \prod_{i \in \mathcal{I}} E_i$, defines $\pi_i : \Omega \longrightarrow E_i$ to be the natural projection map for each $i \in \mathcal{I}$, takes $\mathcal{F}_i = \pi_i^{-1}(\mathcal{B}_i)$, $i \in \mathcal{I}$, and $\mathcal{F} = \bigvee_{i \in \mathcal{I}} \mathcal{F}_i$, and shows that there is a unique probability measure P on $(\Omega, \mathcal{F})$ with the properties that

$$P\big(\pi_i^{-1}\Gamma_i\big) = \mu_i\big(\Gamma_i\big) \quad \text{for all} \quad i \in \mathcal{I} \text{ and } \Gamma_i \in \mathcal{B}_i$$

and the σ-algebras $\mathcal{F}_i$, $i \in \mathcal{I}$, are P-independent. Although this procedure is extremely powerful, it is rather mechanical. For this reason, we have chosen to defer the details of the product construction to Exercise 1.1.14 below and to, instead, spend the rest of this section developing a more hands-on approach to constructing independent sequences of real-valued random variables. Indeed, although the product method is more ubiquitous and has become the construction of choice, the one which we are about to present has the advantage that it shows independent random variables can arise "naturally" and even in a familiar context.

Until further notice, we take $(\Omega, \mathcal{F}) = \big([0,1), \mathcal{B}_{[0,1)}\big)$ (when E is a metric space, we use $\mathcal{B}_E$ to denote the Borel field over E) and P to be the restriction $\lambda_{[0,1)}$ of Lebesgue's measure $\lambda_{\mathbb{R}}$ to $[0,1)$. We next define the **Rademacher functions** R_n, $n \in \mathbb{Z}^+$, on Ω as follows. Define the **integer part** $[t]$ of $t \in \mathbb{R}$ to be the largest integer dominated by t and consider the function $R : \mathbb{R} \longrightarrow \{-1,1\}$ given by

$$R(t) = \begin{cases} -1 & \text{if } \; t - [t] \in \left[0, \frac{1}{2}\right) \\ \;\;1 & \text{if } \; t - [t] \in \left[\frac{1}{2}, 1\right) \end{cases}.$$

The function R_n is then defined on $[0,1)$ by

$$R_n(\omega) = R\big(2^{n-1}\omega\big), \qquad n \in \mathbb{Z}^+ \text{ and } \omega \in [0,1).$$

We will now show that the Rademacher functions are P-independent. To this end, first note that every real-valued function f on $\{-1,1\}$ is of the form $\alpha + \beta x$, $x \in \{-1,1\}$, for some pair of real numbers α and β. Thus, all that we have to show is that

$$\mathbb{E}^P\big[(\alpha_1 + \beta_1 R_1)\cdots(\alpha_n + \beta_n R_n)\big] = \alpha_1 \cdots \alpha_n$$

for any $n \in \mathbb{Z}^+$ and $(\alpha_1, \beta_1), \dots, (\alpha_n, \beta_n) \in \mathbb{R}^2$. Since this is obvious when $n = 1$, we will assume that it holds for n and will deduce that it must also hold for $n+1$; and clearly this comes down to checking that

$$\mathbb{E}^P\big[F(R_1, \dots, R_n)\, R_{n+1}\big] = 0$$

for any $F : \{-1,1\}^n \longrightarrow \mathbb{R}$. But $(R_1, \dots, R_n)$ is constant on each interval

$$I_{m,n} \equiv \left[\frac{m}{2^n}, \frac{m+1}{2^n}\right), \quad 0 \le m < 2^n$$

whereas R_{n+1} integrates to 0 on each $I_{m,n}$. Hence, by writing the integral over Ω as the sum of integrals over the $I_{m,n}$'s, we get the desired result.

At this point we have produced a countably infinite sequence of independent **Bernoulli random variables** (i.e., two-valued random variables whose range is usually either $\{-1,1\}$ or $\{0,1\}$) with mean-value 0. In order to get more general random variables, we combine our Bernoulli random variables together in a clever way.

Recall that a random variable U is said to be **uniformly distributed** on the finite interval $[a,b]$ if

$$P(U \le t) = \frac{t-a}{b-a} \quad \text{for } t \in [a,b].$$

1.1.8 Lemma. *Let $\{Y_\ell : \ell \in \mathbb{Z}^+\}$ be a sequence of P-independent $\{0,1\}$-valued Bernoulli random variables with mean-value $\frac{1}{2}$ on some probability space $(\Omega, \mathcal{F}, P)$, and set*

$$U = \sum_{\ell=1}^{\infty} \frac{X_\ell}{2^\ell}.$$

Then U is uniformly distributed on $[0,1]$.

PROOF: Because the assertion only involves properties of distributions, it will be proved in general as soon as we prove it for a particular realization of independent, mean-value $\frac{1}{2}$, $\{0,1\}$-valued Bernoulli random variables. In particular, by the preceding discussion, we need only consider the random variables

$$\epsilon_n(\omega) \equiv \frac{1 + R_n(\omega)}{2}, \quad n \in \mathbb{Z}^+ \text{ and } \omega \in [0,1),$$

on $\left([0,1), \mathcal{B}_{[0,1)}, \lambda_{[0,1)}\right)$. But, as is easily checked, for each $\omega \in [0,1]$, $\omega = \sum_{n=1}^{\infty} 2^{-n} \epsilon_n(\omega)$. Hence, the desired conclusion is trivial in this case. □

Now let $(k,\ell) \in \mathbb{Z}^+ \times \mathbb{Z}^+ \longmapsto n(k,\ell) \in \mathbb{Z}^+$ be any one-to-one mapping of $\mathbb{Z}^+ \times \mathbb{Z}^+$ onto $\mathbb{Z}^+$, and set

$$Y_{k,\ell} = \frac{1 + R_{n(k,\ell)}}{2}, \qquad (k,\ell) \in \left(\mathbb{Z}^+\right)^2.$$

Clearly, each $Y_{k,\ell}$ is a $\{0,1\}$-valued Bernoulli random variable with mean-value $\frac{1}{2}$, and the family $\left\{Y_{k,\ell} : (k,\ell) \in \left(\mathbb{Z}^+\right)^2\right\}$ is P-independent. Hence, by Lemma 1.1.8, each of the random variables

$$U_k \equiv \sum_{\ell=1}^{\infty} \frac{Y_{k,\ell}}{2^\ell}, \qquad k \in \mathbb{Z}^+,$$

is uniformly distributed on $[0,1)$. In addition, the U_k's are obviously mutually independent. Hence, we have now produced a sequence of mutually independent random variables, each of which is uniformly distributed on $[0,1)$. To complete our program, we use the time-honored transformation which takes a uniform random variable into an arbitrary one. Namely, given a **distribution function** F on $\mathbb{R}$ (i.e., F is a right-continuous, nondecreasing function which tends to 0 at $-\infty$ and 1 at $+\infty$), define F^{-1} on $[0,1]$ to be the left-continuous inverse of F. That is,

$$F^{-1}(t) = \inf\{s \in \mathbb{R} : F(s) \geq t\}, \qquad t \in [0,1].$$

(Throughout, the infimum over the empty set is taken to be $+\infty$.) It is then an easy matter to check that when U is uniformly distributed on $[0,1)$ the random variable $X = F^{-1} \circ U$ has distribution function F:

$$P(X \leq t) = F(t), \qquad t \in \mathbb{R}.$$

Hence, after combining this with what we already know, we have now completed the proof of the following theorem.

1.1.9 Theorem. *Let $\Omega = [0,1)$, $\mathcal{F} = \mathcal{B}_{[0,1)}$, and $P = \lambda_{[0,1)}$. Then for any sequence $\{F_k : k \in \mathbb{Z}^+\}$ of distribution functions on $\mathbb{R}$ there exists a sequence $\{X_k : k \in \mathbb{Z}^+\}$ of P-independent random variables on $(\Omega, \mathcal{F}, P)$ with the property that $P\big(X_k \le t\big) = F_k(t)$, $t \in \mathbb{R}$, for each $k \in \mathbb{Z}^+$.*

Exercises

1.1.10 Exercise: As we pointed out, $P\big(A_1 \cap A_2\big) = P\big(A_1\big)P\big(A_2\big)$ if and only if the σ-algebra generated by A_1 is P-independent of the one generated by A_2. Construct an example to show that the analogous statement is false when dealing with three, instead of two, sets. That is, just because $P\big(A_1 \cap A_2 \cap A_3\big) = P\big(A_1\big)P\big(A_2\big)P\big(A_3\big)$, it is not necessarily true that the three σ-algebras generated by A_1, A_2, and A_3 are P-independent.

1.1.11 Exercise: In this exercise we point out two elementary, but important, properties of independent random variables. Throughout, $(\Omega, \mathcal{F}, P)$ is a given probability space.

(i) Let X_1 and X_2 be a pair of P-independent random variables with values in the measurable spaces $(E_1, \mathcal{B}_1)$ and $(E_2, \mathcal{B}_2)$, respectively. Given a $\mathcal{B}_1 \times \mathcal{B}_2$-measurable function $F : E_1 \times E_2 \longrightarrow \mathbb{R}$ which is either nonnegative or bounded, use Tonelli's or Fubini's Theorem to show that

$$x_2 \in E_2 \longmapsto f(x_2) \equiv \mathbb{E}^P\Big[F\big(X_1, x_2\big)\Big] \in \mathbb{R}$$

is $\mathcal{B}_2$-measurable and that

$$\mathbb{E}^P\Big[F\big(X_1, X_2\big)\Big] = \mathbb{E}^P\Big[f\big(X_2\big)\Big].$$

(ii) Suppose that $X_1, \dots, X_n$ are P-independent, real-valued random variables. If each of the X_m's is P-integrable, show that $X_1 \cdots X_n$ is also P-integrable and that

$$\mathbb{E}^P\big[X_1 \cdots X_n\big] = \mathbb{E}^P\big[X_1\big] \cdots \mathbb{E}^P\big[X_n\big].$$

1.1.12 Exercise: Given a nonempty set Ω, recall[†] that a collection $\mathcal{C}$ of subsets of Ω is called a **π-system** if $\mathcal{C}$ is closed under finite intersections. At the same

[†] See, for example, §3.1 in the author's *A Concise Introduction to the Theory of Integration*, Third Edition publ. by Birkhäuser (1998).

time, recall that a collection $\mathcal{L}$ is called a **λ-system** if $\Omega \in \mathcal{L}$, $A \cup B \in \mathcal{L}$ whenever A and B are disjoint members of $\mathcal{L}$, $B \setminus A \in \mathcal{L}$ whenever A and B are members of $\mathcal{L}$ with $A \subseteq B$, and $\bigcup_1^\infty A_n \in \mathcal{L}$ whenever $\{A_n\}_1^\infty$ is a nondecreasing sequence of members of $\mathcal{L}$. Finally, recall (cf. Lemma 3.1.3 in *ibid.*) that if $\mathcal{C}$ is a π-system, then the σ-algebra $\sigma(\mathcal{C})$ is the smallest $\mathcal{L}$-system $\mathcal{L} \supseteq \mathcal{C}$.

Now, let $(\Omega, \mathcal{F}, P)$ be a probability space, and, for each element i of the index set $\mathcal{I}$, let $\mathcal{C}_i \subseteq \mathcal{F}$ be a π- system. Show that the σ-algebras $\mathcal{F}_i$ generated by the $\mathcal{C}_i$'s are P-independent if and only if (1.1.1) holds for all choices of $n \geq 2$, distinct $i_1, \dots, i_n \in \mathcal{I}$, and $A_{i_1} \in \mathcal{C}_{i_1}, \dots, A_{i_n} \in \mathcal{C}_{i_n}$.

1.1.13 Exercise: In this exercise we discuss two criteria for determining when random variables on the probability space $(\Omega, \mathcal{F}, P)$ are independent.

(i) Let $X_1, \dots,$ and X_n be bounded, real-valued random variables. Using Weierstrass's approximation theorem, show that the X_m's are P-independent if and only if

$$\mathbb{E}^P\left[X_1^{m_1} \cdots X_n^{m_n}\right] = \mathbb{E}^P\left[X_1^{m_1}\right] \cdots \mathbb{E}^P\left[X_n^{m_n}\right]$$

for all $m_1, \dots, m_n \in \mathbb{N}$.

(ii) Let $\mathbf{X} : \Omega \longrightarrow \mathbb{R}^m$ and $\mathbf{Y} : \Omega \longrightarrow \mathbb{R}^n$ be random variables. Show that $\mathbf{X}$ and $\mathbf{Y}$ are P-independent if and only if

$$\begin{aligned}\mathbb{E}^P&\left[\exp\left[\sqrt{-1}\left(\left(\boldsymbol{\alpha}, \mathbf{X}\right)_{\mathbb{R}^m} + \left(\boldsymbol{\beta}, \mathbf{Y}\right)_{\mathbb{R}^n}\right)\right]\right] \\ &= \mathbb{E}^P\left[\exp\left[\sqrt{-1}\left(\boldsymbol{\alpha}, \mathbf{X}\right)_{\mathbb{R}^m}\right]\right] \mathbb{E}^P\left[\exp\left[\sqrt{-1}\left(\boldsymbol{\beta}, \mathbf{Y}\right)_{\mathbb{R}^n}\right]\right]\end{aligned}$$

for all $\boldsymbol{\alpha} \in \mathbb{R}^m$ and $\boldsymbol{\beta} \in \mathbb{R}^n$.

Hint: The *only if* assertion is obvious. To prove the *if* assertion, first check that $\mathbf{X}$ and $\mathbf{Y}$ are independent if

$$\mathbb{E}^P\left[f(\mathbf{X})\, g(\mathbf{Y})\right] = \mathbb{E}^P\left[f(\mathbf{X})\right] \mathbb{E}^P\left[g(\mathbf{Y})\right]$$

for all $f \in C_c^\infty(\mathbb{R}^m; \mathbb{C})$ and $g \in C_c^\infty(\mathbb{R}^n; \mathbb{C})$. Second, given such f and g, apply elementary Fourier analysis to write

$$f(\mathbf{x}) = \int_{\mathbb{R}^m} e^{\sqrt{-1}\,(\boldsymbol{\alpha}, \mathbf{x})_{\mathbb{R}^m}}\, \varphi(\boldsymbol{\alpha})\, d\boldsymbol{\alpha} \quad \text{and} \quad g(\mathbf{y}) = \int_{\mathbb{R}^n} e^{\sqrt{-1}\,(\boldsymbol{\beta}, \mathbf{y})_{\mathbb{R}^n}}\, \psi(\boldsymbol{\beta})\, d\boldsymbol{\beta},$$

where φ and ψ are smooth functions with **rapidly decreasing** (i.e., tending to 0 as $|\mathbf{x}| \to \infty$ faster than any power of $(1 + |\mathbf{x}|)^{-1}$) derivatives of all orders. Finally, apply Fubini's Theorem.

1.1.14 Exercise: Given a pair of measurable spaces $(E_1, \mathcal{B}_1)$ and $(E_2, \mathcal{B}_2)$, recall that their product is the measurable space $\big(E_1 \times E_2, \mathcal{B}_1 \times \mathcal{B}_2\big)$, where $\mathcal{B}_1 \times \mathcal{B}_2$ is the σ-algebra over the Cartesian product space $E_1 \times E_2$ generated by the sets $\Gamma_1 \times \Gamma_2$, $\Gamma_i \in \mathcal{B}_i$. Further, recall that, for any probability measures μ_i on $(E_i, \mathcal{B}_i)$, there is a unique probability measure $\mu_1 \times \mu_2$ on $\big(E_1 \times E_2, \mathcal{B}_1 \times \mathcal{B}_2\big)$ such that

$$(\mu_1 \times \mu_2)\big(\Gamma_1 \times \Gamma_2\big) = \mu_1(\Gamma_1)\mu_2(\Gamma_2) \quad \text{for } \Gamma_i \in \mathcal{B}_i.$$

More generally, for any $n \geq 2$ and measurable spaces $\big\{(E_i, \mathcal{B}_i)\big\}_1^n$, one takes $\prod_1^n \mathcal{B}_i$ to be the σ-algebra over $\prod_1^n E_i$ generated by the sets $\prod_1^n \Gamma_i$, $\Gamma_i \in \mathcal{B}_i$. In particular, since $\prod_1^{n+1} E_i$ and $\prod_1^{n+1} \mathcal{B}_i$ can be identified with $(\prod_1^n E_i) \times E_{n+1}$ and $(\prod_1^n \mathcal{B}_i) \times \mathcal{B}_{n+1}$, respectively, one can use induction to show that, for every choice of probability measures μ_i on $(E_i, \mathcal{B}_i)$, there is a unique probability measure $\prod_1^n \mu_i$ on $(\prod_1^n E_i, \prod_1^n \mathcal{B}_i)$ such that

$$\left(\prod_1^n \mu_i\right)\left(\prod_1^n \Gamma_i\right) = \prod_1^n \mu_i(\Gamma_i), \quad \Gamma_i \in \mathcal{B}_i.$$

The purpose of this exercise is to generalize the preceding construction to infinite collections. Thus, let $\mathfrak{I}$ be an infinite index set, and, for each $i \in \mathfrak{I}$, let $(E_i, \mathcal{B}_i)$ be a measurable space. Given $\emptyset \neq \Lambda \subseteq \mathfrak{I}$, we will use $\mathbf{E}_\Lambda$ to denote the Cartesian product space $\prod_{i \in \Lambda} E_i$ and π_Λ to denote the natural projection map taking $\mathbf{E}_\mathfrak{I}$ onto $\mathbf{E}_\Lambda$. Further, we use $\mathcal{B}_\mathfrak{I} = \prod_{i \in \mathfrak{I}} \mathcal{B}_i$ to stand for the σ-algebra over $\mathbf{E}_\mathfrak{I}$ generated by the collection $\mathcal{C}$ of subsets

$$\pi_F^{-1}\left(\prod_{i \in F} \Gamma_i\right), \quad \Gamma_i \in \mathcal{B}_i,$$

as F varies over nonempty, finite subsets of $\mathfrak{I}$ (abbreviated by: $\emptyset \neq F \subset\subset \mathfrak{I}$). In the following steps, we will outline a proof that, for every choice of probability measures μ_i on $(E_i, \mathcal{B}_i)$, there is a unique probability measure $\prod_{i \in \mathfrak{I}} \mu_i$ on $\big(\mathbf{E}_\mathfrak{I}, \mathcal{B}_\mathfrak{I}\big)$ with the property that

$$\left(\prod_{i \in \mathfrak{I}} \mu_i\right)\left(\pi_F^{-1}\left(\prod_{i \in F} \Gamma_i\right)\right) = \prod_{i \in F} \mu_i(\Gamma_i), \quad \Gamma_i \in \mathcal{B}_i, \tag{1.1.15}$$

for every $\emptyset \neq F \subset\subset \mathfrak{I}$. Not surprisingly, the probability space

$$\left(\prod_{i \in \mathfrak{I}} E_i, \prod_{i \in \mathfrak{I}} \mathcal{B}_i, \prod_{i \in \mathfrak{I}} \mu_i\right)$$

is called the **product** over $\mathfrak{I}$ of the spaces $\big(E_i, \mathcal{B}_i, \mu_i\big)$; and when all the factors are the same space $\big(E, \mathcal{B}, \mu\big)$, it is customary to denote it by $\big(E^\mathfrak{I}, \mathcal{B}^\mathfrak{I}, \mu^\mathfrak{I}\big)$, and if, in addition, $\mathfrak{I} = \{1, \ldots, N\}$, one uses $\big(E^N, \mathcal{B}^N, \mu^N\big)$.

(i) After noting that two probability measures which agree on a π-system agree on the σ-algebra generated by that π-system, show that there is at most one probability measure on $(\mathbf{E}_{\mathfrak{I}}, \mathcal{B}_{\mathfrak{I}})$ which satisfies the condition in (1.1.15). Hence, the problem is purely one of existence.

(ii) Let $\mathcal{A}$ be the *algebra* over $\mathbf{E}_{\mathfrak{I}}$ generated by $\mathcal{C}$, and show that there is a *finitely* additive $\mu : \mathcal{A} \longrightarrow [0,1]$ with the property that

$$\mu\Big(\pi_F^{-1}(\Gamma_F)\Big) = \left(\prod_{i\in F} \mu_i\right)(\Gamma_F), \quad \Gamma_F \in \mathcal{B}_F,$$

for all $\emptyset \neq F \subset\subset \mathfrak{I}$. Hence, all that we have to do is check that μ admits a σ-additive extension to $\mathcal{B}_{\mathfrak{I}}$, and, by Carathéodory's Extension Theorem, this comes down to checking that $\mu(A_n) \searrow 0$ whenever $\{A_n\}_1^\infty \subseteq \mathcal{A}$ and $A_n \searrow \emptyset$. Thus, let $\{A_n\}_1^\infty$ be a nonincreasing sequence from $\mathcal{A}$, and assume that $\mu(A_n) \geq \epsilon$ for some $\epsilon > 0$ and all $n \in \mathbb{Z}^+$. We must show that $\bigcap_1^\infty A_n \neq \emptyset$.

(iii) Referring to the last part of **(ii)**, show that there is no loss in generality if we assume that $A_n = \pi_{F_n}^{-1}(\Gamma_{F_n})$, where, for each $n \in \mathbb{Z}^+$, $\emptyset \neq F_n \subset\subset \mathfrak{I}$ and $\Gamma_{F_n} \in \mathcal{B}_{F_n}$. In addition, show that we may assume that $F_1 = \{i_1\}$ and that $F_n = F_{n-1} \cup \{i_n\}$, $n \geq 2$, where $\{i_n\}_1^\infty$ is a sequence of distinct elements of $\mathfrak{I}$. Now, make these assumptions and show that it suffices for us to find $a_\ell \in E_{i_\ell}$, $\ell \in \mathbb{Z}^+$, with the property, for each $m \in \mathbb{Z}^+$, $(a_1, \dots, a_m) \in \Gamma_{F_m}$.

(iv) Continuing **(iii)**, for each m, $n \in \mathbb{Z}^+$, define $g_{m,n} : \mathbf{E}_{F_m} \longrightarrow [0,1]$ so that

$$g_{m,n}\big(\mathbf{x}_{F_m}\big) = \mathbf{1}_{\Gamma_{F_n}}\big(x_{i_1}, \dots, x_{i_n}\big) \quad \text{if } n \leq m$$

and

$$g_{m,n}\big(\mathbf{x}_{F_m}\big) = \int_{\mathbf{E}_{F_n \setminus F_m}} \mathbf{1}_{\Gamma_{F_n}}\big(\mathbf{x}_{F_m}, \mathbf{y}_{F_n\setminus F_m}\big) \left(\prod_{\ell=m+1}^{n} \mu_{i_\ell}\right) \big(d\mathbf{y}_{F_n\setminus F_m}\big)$$

if $n > m$. After noting that, for each m and n, $g_{m,n+1} \leq g_{m,n}$ and

$$g_{m,n}\big(\mathbf{x}_{F_m}\big) = \int_{E_{i_{m+1}}} g_{m+1,n}\big(\mathbf{x}_{F_m}, y_{i_{m+1}}\big)\, \mu_{i_{m+1}}\big(dy_{i_{m+1}}\big),$$

set $g_m = \lim_{n\to\infty} g_{m,n}$ and conclude that

$$g_m\big(\mathbf{x}_{F_m}\big) = \int_{E_{i_{m+1}}} g_{m+1}\big(\mathbf{x}_{F_m}, y_{i_{m+1}}\big)\, \mu_{i_{m+1}}\big(dy_{i_{m+1}}\big).$$

In addition, note that

$$\begin{aligned}\int_{E_{i_1}} g_1(x_{i_1})\,\mu_{i_1}(dx_{i_1}) &= \lim_{n\to\infty}\int_{E_{i_1}} g_{1,n}(x_{i_1})\,\mu_{i_1}(dx_{i_1})\\ &= \lim_{n\to\infty}\mu(A_n)\ge\epsilon,\end{aligned}$$

and proceed by induction to produce $a_\ell \in E_{i_\ell}$, $\ell\in\mathbb{Z}^+$, so that

$$g_m\big((a_1,\dots,a_m)\big)\ge\epsilon \quad \text{for all } m\in\mathbb{Z}^+.$$

Finally, check that $\{a_m\}_1^\infty$ is a sequence of the sort for which we were looking at the end of part (**iii**).

1.1.16 Exercise: Recall that if Φ is a measurable map from one measurable space $(E,\mathcal{B})$ into a second one $(E',\mathcal{B}')$, then the **distribution** of Φ under a measure μ on $(E,\mathcal{B})$ is the **pushforward** measure $\Phi_*\mu$ (also denoted by $\mu\circ\Phi^{-1}$) defined on $(E',\mathcal{B}')$ by

$$\Phi_*\mu(\Gamma)=\mu\big(\Phi^{-1}(\Gamma)\big)\quad\text{for}\quad \Gamma\in\mathcal{B}'.$$

Given a nonempty index set $\mathfrak{I}$ and, for each $i\in\mathfrak{I}$, a measurable space $(E_i,\mathcal{B}_i)$ and an E_i-valued random variable X_i on the probability space $(\Omega,\mathcal{F},P)$, define $\mathbf{X}:\Omega\longrightarrow\prod_{i\in\mathfrak{I}}E_i$ so that $\mathbf{X}(\omega)_i = X_i(\omega)$ for each $i\in\mathfrak{I}$ and $\omega\in\Omega$. Show that $\{X_i : i\in\mathfrak{I}\}$ is a family of P-independent random variables if and only if $\mathbf{X}_*P=\prod_{i\in\mathfrak{I}}(X_i)_*P$. In particular, given probability measures μ_i on $(E_i,\mathcal{B}_i)$, set

$$\Omega=\prod_{i\in\mathfrak{I}}E_i,\quad \mathcal{F}=\prod_{i\in\mathfrak{I}}\mathcal{B}_i,\quad P=\prod_{i\in\mathfrak{I}}\mu_i,$$

let $X_i:\Omega\longrightarrow E_i$ be the natural projection map from Ω onto E_i, and show that $\{X_i : i\in\mathfrak{I}\}$ is a family of mutually P-independent random variables such that, for each $i\in\mathfrak{I}$, X_i has distribution μ_i.

1.1.17 Exercise: Although it does not entail infinite product spaces, an interesting example of the way in which the preceding type of construction can be effectively applied is provided by the following elementary version of a *coupling* argument.

(**i**) Let $(\Omega,\mathcal{B},P)$ be a probability space and X and Y a pair of square P-integrable $\mathbb{R}$-valued random variables with the property that

$$\big(X(\omega)-X(\omega')\big)\,\big(Y(\omega)-Y(\omega')\big)\ge 0\quad\text{for all } (\omega,\omega')\in\Omega^2.$$

Show that

$$\mathbb{E}^P\big[X\,Y\big] \geq \mathbb{E}^P[X]\,\mathbb{E}^P[Y].$$

Hint: Define X_i and Y_i on Ω^2 for $i \in \{1,2\}$ so that $X_i(\boldsymbol{\omega}) = X(\omega_i)$ and $Y_i(\boldsymbol{\omega}) = Y(\omega_i)$ when $\boldsymbol{\omega} = (\omega_1, \omega_2)$, and integrate the inequality

$$0 \leq \big(X(\omega_1) - X(\omega_2)\big)\,\big(Y(\omega_1) - Y(\omega_2)\big) = \big(X_1(\boldsymbol{\omega}) - X_2(\boldsymbol{\omega})\big)\,\big(Y_1(\boldsymbol{\omega}) - Y_2(\boldsymbol{\omega})\big)$$

with respect to P^2.

(ii) Suppose that $n \in \mathbb{Z}^+$ and that f and g are $\mathbb{R}$-valued, Borel measurable functions on $\mathbb{R}^n$ which are nondecreasing with respect to each coordinate (separately). Show that if $\mathbf{X} = \big(X_1, \dots, X_n\big)$ is an $\mathbb{R}^n$-valued random variable on a probability space $(\Omega, \mathcal{B}, P)$ whose coordinates are mutually P-independent, then

$$\mathbb{E}^P\big[f(\mathbf{X})\,g(\mathbf{X})\big] \geq \mathbb{E}^P\big[f(\mathbf{X})\big]\,\mathbb{E}^P\big[g(\mathbf{X})\big]$$

so long as $f(\mathbf{X})$ and $g(\mathbf{X})$ are both square P-integrable.

Hint: First check that the case when $n = 1$ reduces to an application of (**i**). Next, describe the general case in terms of a multiple integral, apply Fubini's Theorem, and make repeated use of the case when $n = 1$.

1.1.18 Exercise: A σ-algebra is said to be **countably generated** if it contains a countable collection of sets which generate it. In this exercise, we will show that just because a σ-algebra is itself countably generated does not mean that all its sub-σ-algebras are.

Let $(\Omega, \mathcal{F}, P)$ be a measurable space and $\big\{\mathcal{F}_n : n \in \mathbb{Z}^+\big\}$ be a sequence of P-independent sub-σ-algebras of $\mathcal{F}$. Further, assume that, for each $n \in \mathbb{Z}^+$, there is an $A_n \in \mathcal{F}_n$ which satisfies $\alpha \leq P\big(A_n\big) \leq 1-\alpha$ for some fixed $\alpha \in (0, \frac{1}{2})$. Show that the tail σ-algebra $\mathcal{T}$ determined by $\big\{\mathcal{F}_n : n \in \mathbb{Z}^+\big\}$ cannot be countably generated.

Hint: First, reduce to the case when each $\mathcal{F}_n$ is generated by the set A_n. After making this reduction, show that C is an **atom** in $\mathcal{T}$ (i.e., $B = C$ whenever $B \in \mathcal{T} \setminus \{\emptyset\}$ is contained in C) only if one can write

$$C = \varliminf_{n\to\infty} C_n \equiv \bigcup_{m=1}^{\infty} \bigcap_{n\geq m} C_n$$

where, for each $n \in \mathbb{Z}^+$, either C_n equals A_n or $A_n\complement$. Conclude that every atom in $\mathcal{T}$ must have P-measure 0. Now suppose that $\mathcal{T}$ were generated by $\big\{B_\ell : \ell \in \mathbb{N}\big\}$. By Kolmogorov's 0–1 Law (cf. Theorem 1.1.2), $P\big(B_\ell\big) \in \{0,1\}$ for every $\ell \in \mathbb{N}$. Take

$$\hat{B}_\ell = \begin{cases} B_\ell & \text{if } \; P\big(B_\ell\big) = 1 \\ B_\ell\complement & \text{if } \; P\big(B_\ell\big) = 0 \end{cases} \quad \text{and set} \quad C = \bigcap_{\ell\in\mathbb{N}} \hat{B}_\ell.$$

Note that, on the one hand, $P(C) = 1$, while, on the other hand, C is an atom in $\mathcal{T}$.

1.1.19 Exercise: Here is an application of Kolmogorov's 0–1 Law to Lebesgue's measure on $[0, 1)$.

(**i**) Referring to the discussion preceding Lemma 1.1.8, define the transformations $T_n : [0, 1) \longrightarrow [0, 1)$ for $n \in \mathbb{Z}^+$ so that

$$T_n(\omega) = \omega - \frac{1 + R_n(\omega)}{2^{n+1}}, \quad \omega \in [0, 1),$$

and notice (cf. the proof of Lemma 1.1.8) that $T_n(\omega)$ simply *flips* the nth coefficient in the binary expansion ω. Next, let $\Gamma \in \mathcal{B}_{[0,1)}$, and show that Γ is measurable with respect of the σ-algebra $\sigma\big(R_n : n > m\big)$ generated by $\{R_n : n > m\}$ if and only if $T_n(\Gamma) = \Gamma$ for each $1 \le n \le m$. In particular, conclude that $\lambda_{[0,1)}(\Gamma) \in \{0, 1\}$ if $T_n\Gamma = \Gamma$ for every $n \in \mathbb{Z}^+$.

(**ii**) Let $\mathfrak{F}$ denote the set of all finite subsets of $\mathbb{Z}^+$, and for each $F \in \mathfrak{F}$, define $T^F : [0, 1) \longrightarrow [0, 1)$ so that $T^\emptyset$ is the identity mapping and

$$T^{F \cup \{m\}} = T^F \circ T_m \quad \text{for each } F \in \mathfrak{F} \text{ and } m \in \mathbb{Z}^+ \setminus F.$$

As an application of (**i**), show that for every $\Gamma \in \mathcal{B}_{[0,1)}$ with $\lambda_{[0,1)}(\Gamma) > 0$,

$$\lambda_{[0,1)}\left(\bigcup_{F \in \mathfrak{F}} T^F(\Gamma)\right) = 1.$$

In particular, this means that if Γ has positive measure, then almost every $\omega \in [0, 1)$ can be moved to Γ by *flipping* a finite number of the coefficients in the binary expansion of ω.

§1.2: The Weak Law of Large Numbers

Starting with this section, and for the rest of this chapter, we will be studying what happens when one averages P-independent, real-valued random variables. The remarkable fact, which will be confirmed repeatedly, is that the limiting behavior of such averages depends hardly at all on the variables involved. Intuitively, one can explain this phenomenon by pretending that the random variables are building blocks which, in the averaging process, first get homothetically shrunk and then reassembled according to a regular pattern. Hence, by the time that one passes to the limit, the peculiarities of the original blocks get lost.

Throughout our discussion, $(\Omega, \mathcal{F}, P)$ will be a probability space on which we have a sequence $\{X_n\}_1^\infty$ of real-valued random variables. Given $n \in \mathbb{Z}^+$, we will use S_n to denote the partial sum $X_1 + \cdots + X_n$ and $\overline{S}_n$ to denote the average

$$\frac{S_n}{n} = \frac{1}{n}\sum_{\ell=1}^{n} X_\ell.$$

Our first result is a very general one; in fact, it even applies to random variables which are not necessarily independent and do not necessarily have mean 0.

1.2.1 Lemma. *Assume that*

$$\mathbb{E}^P\left[X_n^2\right] < \infty \text{ for } n \in \mathbb{Z}^+ \quad \text{and} \quad \mathbb{E}^P\left[X_k X_\ell\right] = 0 \text{ if } k \neq \ell. \tag{1.2.2}$$

Then, for each $\epsilon > 0$,

$$\epsilon^2 P\Big(\left|\overline{S}_n\right| \geq \epsilon\Big) \leq \mathbb{E}^P\left[\overline{S}_n^2\right] = \frac{1}{n^2}\sum_{\ell=1}^{n} \mathbb{E}^P\left[X_\ell^2\right] \quad \text{for } n \in \mathbb{Z}^+. \tag{1.2.3}$$

In particular, if

$$M \equiv \sup_{n \in \mathbb{Z}^+} \mathbb{E}^P\left[X_n^2\right] < \infty, \tag{1.2.4}$$

then

$$\epsilon^2 P\Big(\left|\overline{S}_n\right| \geq \epsilon\Big) \leq \mathbb{E}^P\left[\overline{S}_n^2\right] \leq \frac{M}{n}, \qquad n \in \mathbb{Z}^+ \text{ and } \epsilon > 0; \tag{1.2.5}$$

and so $\overline{S}_n \longrightarrow 0$ in $L^2(P)$ and also in P-probability.

Proof: To prove the equality in (1.2.3), note that, by (1.2.2),

$$\mathbb{E}^P\left[S_n^2\right] = \sum_{\ell=1}^{n} \mathbb{E}^P\left[X_\ell^2\right].$$

The rest is just an application of **Chebyshev's inequality**, the estimate which results after integrating the inequality

$$\epsilon^2 \mathbf{1}_{[\epsilon,\infty)}\big(|Y|\big) \leq Y^2 \mathbf{1}_{[\epsilon,\infty)}\big(|Y|\big) \leq Y^2$$

for any random variable Y. $\square$

Obviously, Lemma 1.2.1 has less to do with the property of independence than it does with Bessel's inequality for general orthogonal functions. On the other hand, independent random variables provide a ready source of orthogonal functions. Indeed, recall that for any P-integrable random variable X, its **variance** $\operatorname{var}(X)$ satisfies

$$\operatorname{var}(X) \equiv \mathbb{E}^P\left[\left(X - \mathbb{E}^P[X]\right)^2\right] = \mathbb{E}^P\left[X^2\right] - \left(\mathbb{E}^P[X]\right)^2 \leq \mathbb{E}^P\left[X^2\right]. \tag{1.2.6}$$

In particular, if the random variables X_n, $n \in \mathbb{Z}^+$, are P-independent and satisfy the first part of (1.2.2), then the random variables

$$\hat{X}_n \equiv X_n - \mathbb{E}^P\left[X_n\right] \qquad n \in \mathbb{Z}^+,$$

are still square P-integrable, now have mean-value 0, and therefore satisfy the whole of (1.2.2). Hence, the following statement is an immediate consequence of Lemma 1.2.1.

1.2.7 Theorem. *Let $\{X_n : n \in \mathbb{Z}^+\}$ be a sequence of P-independent, square P-integrable random variables with mean-value m and variance dominated by σ^2. Then, for every $n \in \mathbb{Z}^+$ and $\epsilon > 0$:*

$$\epsilon^2\, P\left(\left|\overline{S}_n - m\right| \geq \epsilon\right) \leq \mathbb{E}^P\left[\left(\overline{S}_n - m\right)^2\right] \leq \frac{\sigma^2}{n}. \tag{1.2.8}$$

In particular, $\overline{S}_n \longrightarrow m$ in $L^2(P)$ and therefore in P-probability.

As yet we have only made minimal use of independence: all that we have done is subtract off the mean of independent random variables and thereby made them orthogonal. In order to bring the full force of independence into play, one has to exploit the fact that one can compose independent random variables with any (measurable) functions without destroying their independence; in particular, *truncating* independent random variables does not destroy independence. To see how such a property can be brought to bear, we will now consider the problem of extending the last part of Theorem 1.2.7 to X_n's which are less than square P-integrable. In order to understand the statement, recall that a family $\{X_i : i \in \mathcal{I}\}$ of random variables is said to be **uniformly P-integrable** if

$$\lim_{R \nearrow \infty} \sup_{i \in \mathcal{I}} \mathbb{E}^P\left[\left|X_i\right|,\, \left|X_i\right| \geq R\right] = 0. \tag{1.2.9}$$

As the proof of the following theorem illustrates, the importance of this condition is that it allows one to simultaneously approximate the random variables X_i, $i \in \mathcal{I}$, by bounded random variables.

1.2.10 Theorem (The Weak Law of Large Numbers). *Let $\{X_n : n \in \mathbb{Z}^+\}$ be a uniformly P-integrable sequence of P-independent random variables. Then*

$$\frac{1}{n}\sum_1^n \left(X_m - \mathbb{E}^P[X_m]\right) \longrightarrow 0 \text{ in } L^1(P)$$

and, therefore, also in P-probability. In particular, if $\{X_n : n \in \mathbb{Z}^+\}$ is a sequence of P-independent, P-integrable random variables which are identically distributed, then $\overline{S}_n \longrightarrow \mathbb{E}^P[X_1]$ in $L^1(P)$ and P-probability. (Cf. Exercise 1.2.15 below.)

PROOF: Without loss in generality, we will assume that $\mathbb{E}^P[X_n] = 0$ for every $n \in \mathbb{Z}^+$.

For each $R \in (0,\infty)$, define $f_R(t) = t\,\mathbf{1}_{[-R,R]}(t)$, $t \in \mathbb{R}$,

$$m_n^{(R)} = \mathbb{E}^P\left[f_R \circ X_n\right], \quad X_n^{(R)} = f_R \circ X_n - m_n^{(R)}, \quad \text{and} \quad Y_n^{(R)} = X_n - X_n^{(R)},$$

and set

$$\overline{S}_n^{(R)} = \frac{1}{n}\sum_{\ell=1}^n X_\ell^{(R)} \quad \text{and} \quad \overline{T}_n^{(R)} = \frac{1}{n}\sum_{\ell=1}^n Y_\ell^{(R)}.$$

Since $\mathbb{E}[X_n] = 0 \implies m_n^{(R)} = -\mathbb{E}\big[X_n,\, |X_n| > R\big]$,

$$\begin{aligned}\mathbb{E}^P\left[\left|\overline{S}_n\right|\right] &\le \mathbb{E}^P\left[\left|\overline{S}_n^{(R)}\right|\right] + \mathbb{E}^P\left[\left|\overline{T}_n^{(R)}\right|\right] \\ &\le \mathbb{E}^P\left[\left|\overline{S}_n^{(R)}\right|^2\right]^{\frac{1}{2}} + 2\max_{1\le\ell\le n}\mathbb{E}^P\left[|X_\ell|,\, |X_\ell| \ge R\right] \\ &\le \frac{R}{\sqrt{n}} + 2\max_{\ell\in\mathbb{Z}^+}\mathbb{E}^P\left[|X_\ell|,\, |X_\ell| \ge R\right];\end{aligned}$$

and therefore, for each $R > 0$,

$$\varlimsup_{n\to\infty}\mathbb{E}^P\left[\left|\overline{S}_n\right|\right] \le 2\sup_{\ell\in\mathbb{Z}}\mathbb{E}^P\left[|X_\ell|,\, |X_\ell| \ge R\right].$$

Hence, because the X_ℓ's are uniformly P-integrable, we get the desired convergence in $L^1(P)$ by letting $R \nearrow \infty$. □

The name of Theorem 1.2.10 comes from a somewhat invidious comparison with the result in Theorem 1.4.11. The reason why the appellation *weak* is not entirely fair is that, although The Weak Law is indeed less *refined* than the result in Theorem 1.4.11, it is every bit as useful as the one in Theorem 1.4.11 and maybe even more important when it comes to applications. Indeed, what The Weak Law does is provide us with a ubiquitous technique for constructing an **approximate identity** (i.e., a sequence of measures which approximate a point mass) and measuring how fast the approximation is taking place. To illustrate how clever selection of the random variables entering The Weak Law can lead to interesting applications, we will spend the rest of this section discussing S. Bernstein's approach to Weierstrass's approximation theorem.

For a given $p \in [0,1]$, let $\{X_n : n \in \mathbb{Z}^+\}$ be a sequence of P-independent $\{0,1\}$-valued Bernoulli random variables with mean-value p. Then

$$P(S_n = \ell) = \binom{n}{\ell} p^\ell (1-p)^{n-\ell} \quad \text{for} \quad 0 \le \ell \le n.$$

Hence, for any $f \in C([0,1];\mathbb{R})$, the nth **Bernstein polynomial**

$$B_n(p;f) \equiv \sum_{\ell=0}^{n} \binom{n}{\ell} f\left(\frac{\ell}{n}\right) p^\ell (1-p)^{n-\ell} \tag{1.2.11}$$

of f at p is equal to

$$\mathbb{E}^P\left[f \circ \overline{S}_n\right].$$

In particular,

$$\begin{aligned}\left|f(p) - B_n(p;f)\right| &= \left|\mathbb{E}^P\left[f(p) - f \circ \overline{S}_n\right]\right| \le \mathbb{E}^P\left[\left|f(p) - f \circ \overline{S}_n\right|\right] \\ &\le 2\|f\|_{\mathrm{u}} P\left(\left|\overline{S}_n - p\right| \ge \epsilon\right) + \rho(\epsilon;f),\end{aligned}$$

where $\|f\|_{\mathrm{u}}$ is the **uniform norm** of f (i.e., the supremum of $|f|$ over the domain of f) and

$$\rho(\epsilon;f) \equiv \sup\left\{|f(t) - f(s)| : 0 \le s < t \le 1 \text{ with } t - s \le \epsilon\right\}$$

is the modulus of continuity of f. Noting that $\mathrm{var}(X_n) = p(1-p) \le \frac{1}{4}$ and applying (1.2.8), we conclude that, for every $\epsilon > 0$,

$$\left\|f(p) - B_n(p;f)\right\|_{\mathrm{u}} \le \frac{\|f\|_{\mathrm{u}}}{2n\epsilon^2} + \rho(\epsilon;f)$$

In other words, for all $n \in \mathbb{Z}^+$,

$$\left\|f - B_n(\,\cdot\,;f)\right\|_{\mathrm{u}} \le \beta(n;f) \equiv \inf\left\{\frac{\|f\|_{\mathrm{u}}}{2n\epsilon^2} + \rho(\epsilon;f) : \epsilon > 0\right\}. \tag{1.2.12}$$

Obviously, (1.2.12) not only shows that, as $n \to \infty$, $B_n(\,\cdot\,;f) \longrightarrow f$ uniformly on $[0,1]$, but it even provides a rate of convergence in terms of the modulus of continuity of f. Thus, we have done more than simply prove Weierstrass's theorem; we have produced a rather explicit and tractable sequence of approximating polynomials, the sequence $\{B_n(\,\cdot\,;f) : n \in \mathbb{Z}^+\}$. Although this sequence is, by no means, the most efficient one,† as we are about to see, the Bernstein polynomials have a lot to recommend them. In particular, they have the feature that

† See G.G. Lorentz's *Bernstein Polynomials*, Chelsea Publ. Co., New York (1986) for a lot more information.